DOCUMENT RESUME

ED 222 520 TM 820 666

AUTHOR Cranton, Patricia A.
TITLE Statistical Analyses of Evaluation Results.
PUB DATE Mar 82
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Students; Course Evaluation; Educational Assessment; Higher Education; *Outcomes of Education; *Program Evaluation; *Questionnaires; *Statistical Analysis; Student Attitudes; *Student Evaluation of Teacher Performance

ABSTRACT

Student questionnaire data concerning the evaluation of instruction were collected over a 3-day period. Multivariate analyses of variance were used to investigate the statistical analyses of evaluation data. It was found that: departments differed significantly on overall ratings of instruction; class size and level of instruction affected all overall ratings except those of instructor effectiveness; the effect of class size and level of instruction varied dramatically across departments; and ratings of teaching skills were different across departments, levels of instruction, and class sizes (within departments these relationships varied further). The interest/atmosphere factor was the best predictor of overall ratings of instruction. When this relationship was examined in different subgroupings, it was found that lower level students in general placed more value on organization and clarity. It is possible to pinpoint courses which are generally rated lower or higher than others, and this does not seem to be dependent on level: both the highest and the lowest ratings were most often received by level 4 courses. The generation of questions which could be answered from analyses of questionnaire data is also discussed. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED222520

BEST COPY AVAILABLE

STATISTICAL ANALYSES OF EVALUATION RESULTS

Patricia A. Cranton

Centre for Teaching and Learning Services

McGill University

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. A. Cranton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 666

Paper presented at the Annual Meeting of the American Educational Research Association, New York, 1982.

With the current widespread use of student questionnaires in the evaluation of instruction, it is often the case that departments or faculties cumulate large files of questionnaire results. Such results contain a variety of potentially useful information about the nature of the instruction in a particular faculty or department, and perhaps generalizable information about the nature of student ratings.

The current project consisted of three parts: (1) the generation of questions which could be answered from analyses of questionnaire data; (2) statistical analyses of such data; and (3) a discussion of ways in which the results obtained could be used by a university administrator to improve the quality of instruction. This paper is concerned specifically with the second part, the statistical analyses of evaluation data.

Data were collected in one faculty using a modification of the "Teaching Analysis by Students " (TABS) questionnaire developed at the Clinic to Improve University Teaching, University of Massachusetts. Approximately 55,000 students in 1,777 classes responded to the questionnaire over three years (six semesters).

QUESTIONS GENERATED

Questions that could be answered from analyses of these data were:

1. Do student responses to items related to the overall quality of instruction vary dependent on such characteristics as size of the class, level of the course, subject area (department), full time versus part time instructors, day or evening courses, or the campus at which the courses were taught?

2. Aside from the issue of overall quality, are there particular

teaching skills, or clusters of skills which vary dependent on the characteristics of the courses, as listed in question 1?

3. Are student ratings of particular teaching skills more likely to predict student ratings of the overall quality of instruction, and do these relationships vary among the course characteristics which were shown to affect either overall or specific skills ratings in questions 1 and 2?

4. In making personnel decisions, what factors should be considered in the interpretation of evaluation results; e.g., with whom would an individual professor be compared; how should changes in student ratings over time be viewed?

5. In making curriculum decisions, can we identify specific courses within a department which are rated substantially lower, and therefore may be candidates for revision?

ANALYSIS

Several analyses were performed for each question.

1. For each of the characteristics listed in the question, a multivariate analysis of variance was performed with the class mean ratings of four overall items being the dependent variables, and the particular factor under investigation (e.g., full-time/part-time) being the independent variable. Since the four overall ratings are correlated, they must be analyzed simultaneously in order to take into account the correlations among them. Where interactions were expected to exist among factors (level and class size; department and level), a two factor design, with the same four dependent variables, was used. For the continuous variable, class size, four categories were used: sizes of 1 to 20, 21 to 40, 41 to 99, and over 100. This

categorization was based in part on the frequency distribution of class size, and in part on the effect of various class sizes on teaching methods or styles.

2. (a) For each of the course characteristics listed, a multivariate analysis of variance was performed with the class mean ratings of twenty questionnaire items related to specific teaching skills being the dependent variables, and the particular factor under investigation being the dependent variable. (b) In order to investigate whether or not clusters of teaching skills varied across course characteristics, the correlations among class mean ratings were first factor analyzed using a principal components analysis. Factor scores were estimated for each of the four factors obtained, and the multivariate analyses of variance were repeated with the factor scores being the dependent variables, and the same factors as above being the independent variables.

3. (a) Four linear multivariate regression analyses were performed with each of the four overall ratings of the quality of instruction being a dependent variable in one analysis, and with the four factors being the set of independent variables. (b) Based on the results obtained in the investigation of questions 1 and 2, the regression analyses were repeated for the subgroups which were shown to vary on overall ratings and specific skills ratings.

4. Results of the analyses conducted for the previous questions can be interpreted to answer this question. For example, significant differences among various subgroups should be considered in making comparisons among professors. Although not reported in this paper, it would also be useful to investigate changes in ratings over time for individual professors. A two-way multivariate analysis of variance could be conducted with the four overall

ratings of instruction being the dependent variables, and with individual professors being levels of the first factor and times being levels of the second factor. Using professors as one factor, rather than performing separate analyses for individual professors would also allow the examination of the interaction between professors and time, i.e., do some professors change over time while others do not?

5. In order to compare courses, a one-way multivariate analysis of variance was conducted within each department. All offerings of the same course number were grouped together, and were used as levels of the one factor (independent variable). The four overall ratings of instruction were used as dependent variables.

SUMMARY OF RESULTS

I. Overall Quality

The first question addressed was whether or not there were overall differences in student ratings among various course characteristics. In order to answer this question multivariate analyses of variance were used, with four overall student ratings being four dependent variables. The overall ratings were: (1) how much do you think you're learning in this course compared to other courses; (2) how significant is what you're being asked to learn in this course; (3) in comparison to other instructors you have had, how would you rate the effectiveness of the instructor, and (4) how would you rate the overall value of the course.

Due to the large sample size, the level of significance for the tests was set at $p > .0001$.

Full Time Versus Part-Time Instructors

No overall significant differences were found between full-time and part-time instructors. Therefore no further analyses were performed.

Campus

No overall significant differences were found between the two campuses included in the investigation. Campus was not considered further.

Day Versus Evening Courses

A significant overall multivariate difference was found between day and evening courses. However when each of the four dependent variables was tested separately no significant differences between day and evening courses were found. This indicates that a difference exists in the pattern among the four ratings between day and evening courses; i.e., items are correlated differently, but there are no significant differences between means.

Department (Subject Area)

An overall significant multivariate difference was found among the five departments. Further analyses revealed that there were significant differences among departments on each of the four ratings, although these differences were not necessarily of the same magnitude or in the same direction for each item. On the question of amount learned departments 1, 2 and 3 received similar ratings, with departments 4 and 5 being lower, but similar to each other. When ratings of the significance of learning were considered, each department varied from the others, department 1 receiving the highest rating and department 5 the lowest with the remaining departments falling in sequential order between them (see Figure 1). For

the rating of teaching effectiveness, department 3 received the highest ratings, and department 4 the lowest, with department 1, 2 and 5 falling in between and not differing from one another. Overall value of the course produced differences that appear to be a merging of the previous items: departments 1, 2 and 3 are similar, and higher than departments 4 and 5.

The differences among departments clearly have implications for the analyses of other course characteristics. For example, any overall differences in ratings found among levels of instruction or class sizes will have to be considered separately for each department.

Level of Instruction

A significant overall multivariate difference was found among the levels of instruction. When the univariate differences were explored, differences were found among levels on the ratings of amount learned, significance of the learning, and the overall value of the course but not on ratings of instructor effectiveness. The trend of the differences followed that usually found in the research, i.e., higher levels of instruction received higher ratings on the items listed above (see Figure 2). It is important however, that these differences be investigated in each department since department differences were also found. These analyses will be presented in a later section of the paper.

Class Size

Again, a significant overall multivariate difference was found among the categories of class size (1-20, 21-40, 41-100, over 100). When the four items were examined separately, differences among class size were found on ratings of amount learned, significance of learning, and overall value of the

course (Figure 3). As with level of instruction, no significant differences were found for the rating of teacher effectiveness. It was noted that the variance of ratings of teacher effectiveness was larger than that of the other items which may contribute to the lack of significance. Also, it is necessary to further investigate these differences within departments and within levels of instruction since these variables also exhibited different ratings.

Further "Overall Quality" Analyses

Ratings on the overall quality of instruction were found to be significantly different among departments, levels of instruction, and class sizes. If this information is to be of use to administrators, it is necessary to further pinpoint where the differences lie, i.e., are all levels within a specific department different? Do different class sizes within one level of instruction receive different ratings, or do class sizes only produce different ratings because there tend to be larger classes at lower levels? In order to answer questions of this type, three types of analyses were conducted: (1) analyses of the differences among levels within each department, (2) analyses of the differences among class sizes for each level of instruction, and (3) analyses of the differences among class sizes within each department. Again multivariate analyses of variance were performed with the four overall ratings being the four dependent variables.

Level of instruction within each department. In general, the differences among levels varied with the department being investigated. In department 1, the level 3 courses were rated higher on significance of learning and the overall value of the course; no other differences were found. When departments were being compared generally, department 1 was

rated highly on these items; perhaps the level 3 courses are contributing most to this overall difference. In department 2, no differences in ratings among levels were found. In department 3, the general trend was repeated; that is, differences were found among all the levels, with the higher level courses receiving higher ratings. These differences did not exist for the ratings of instructor effectiveness, which was the case for the overall analysis. In department 4, an overall multivariate difference was found, but no differences were found when the individual items were examined. That is, the pattern among item ratings varies with the level, but there are no mean differences; the higher levels of instruction do not receive higher ratings. The violation of the expected pattern occurs at levels 3 and 4. Department 5 also contained an overall significant multivariate difference among levels. Further analysis showed that this difference occurred for ratings of teaching effectiveness only, and it was the fourth level which contributed most to this difference. This was the only department in which this item varied across levels. For other items in this department the pattern of ratings varied across levels, but no mean differences were found. It is interesting to note that department 5 was rated lowest overall on instructor effectiveness; this difference may be largely due to level 4 courses.

Although a general trend exists for higher level courses to receive higher ratings, administrators must view this "rule" with caution. In some departments, this relationship does not exist at all, and in other departments, it may be contributed to by only some of the levels of instruction.

Class size within each level of instruction. It was found previously that ratings decreased as class size increased. However since ratings also tended to increase as the level of instruction increased (where smaller

classes occur more frequently) it was necessary to examine how class size influenced ratings within each level. At the 200 level of instruction (level 2), an overall significant multivariate difference was found. When individual items were examined, it was found that this difference existed only for the ratings of the significance of the learning. That is, at the 200 level, smaller classes yielded higher ratings on the significance of what was learned. Ratings on the instructors' effectiveness showed the least difference due to size within this level. At each of the other levels, 300 level, 400 level, and 600 level, no significant differences were found among class sizes. That is, when one specific level is examined, the overall ratings are not dependent on the size of the class. However, it must be remembered that ratings do vary with the level of the instruction, although it is not consistent across departments.

Class size within each department. Since overall ratings of instruction varied across departments, it was then necessary to examine the relationship between ratings and class size within each department. As for previous analyses, it was found that the effect of class size was not consistent in each department. In departments 1 and 2 there were no significant multivariate differences among the class sizes. Department 3 did show a significant multivariate difference. Further analyses indicated that this difference existed only for ratings of the significance of what was learned, with larger classes having consistently lower ratings. The largest category (size over 100) yielded the most dramatic drop in ratings. Similar results were found in department 4, that is, a significant multivariate difference was found, and when items were examined individually the difference was found to be on the ratings of the significance of what was learned. However in this case, it was the class size of 41 to 100 which was given the lowest

ratings. In department 5 no significant differences were found among class sizes.

In summary, although there may be an overall tendency for student ratings to decrease with class size, this relationship does not hold true in all situations. In some departments, or subject areas, there are no significant differences in ratings among class sizes, and the direct relationship of larger classes yielding lower ratings does not necessarily hold. Again, the administrator who is interpreting ratings in the light of course characteristics such as class size must do so with caution.

II. Teaching Skills

Variations were found, in question 1 across departments, levels of instruction and class sizes on students' ratings of the overall quality of instruction. Next, the question asked was whether differences would also be found when specific teaching skills were considered. The questionnaire contained twenty items which were evaluating, for example, organization of instruction, discussion leading, asking questions and providing feedback to students. In the first set of analyses, these twenty items were used as a set of twenty dependent variables and multivariate analyses of variance were performed with department, level and class size being the independent variables. Overall statistically significant multivariate differences were found for each of the independent variables. In addition, all univariate differences (for the twenty items) were statistically significant. Since this does not provide useful information to an administrator, it was decided to reduce the data to clusters of skills, or factors, and to then investigate the differences among departments, levels and class sizes on these clusters.

Factor Analysis

A principal components analysis of the correlations among the twenty items, with class means being the unit of analysis, yielded four factors (after rotation) which were labelled interest/atmosphere, organization and clarity, evaluation, and discussion. These factors accounted for 83.1%, 8.9%, 4.4% and 3.6% of the variance respectively. Factor scores were then estimated for each mean rating. Rather than twenty skill ratings, the data now consisted of four factor scores which represented ratings on clusters of skills. The multivariate analyses of variance were then repeated using the factor scores as a set of four dependent variables.

Department

An overall significant multivariate difference was found across departments. Further analyses revealed that the departments differed on the interest/atmosphere factor, the evaluation factor and the discussion factor. No differences were found on organization and clarity.

For the interest/atmosphere factor, the largest difference was found between department 4 (high rating) and department 5. This result is consistent with the differences among departments which were found on overall ratings of teaching effectiveness. For the evaluation factor, the largest difference was found between department 4 and department 5; however it is interesting that department 4 received the lowest rating on this factor. Although department 4 was rated highly in terms of overall teaching effectiveness and department 5 was rated low, students rate the evaluation skill (making clear how performance is to be evaluated, evaluation procedures consistent with course goals, etc.) low in department 4 and high in department 5. On the discussion factor, department 4 was rated highly

and department 3 and 5 low. These are expected results in terms of the subject areas; for example courses in department 5 are quantitative in nature and the use of discussion is not likely to be as frequent.

In summary, departments vary in the way in which clusters of skills are rated by students. As was found with ratings of overall quality, the nature of these differences is complex, and departments which are rated highly on one cluster of skills are not necessarily rated highly on another. An administrator who wished to improve the overall quality of instruction would want to focus the effort on different aspects of teaching in different subject areas.

Level

An overall significant multivariate difference on the four factors was found across the levels of instruction. When the factors were examined separately, differences were found on interest/atmosphere, evaluation, and discussion. Again, there were no differences in organization and clarity among levels.

On the interest/atmosphere factor, level 4 was rated the lowest and level 6 the highest. This varies from the expected trend of ratings increasing as level increases; it might be speculated that students who are in their graduating year (level 4) have different expectations of their courses. On the evaluation factor, level 4 courses were also rated the lowest, followed by level 6, and level 2 courses were rated the highest. Again the expected trend is not found and this may be related to the expectations of graduating and graduate (level 6) students. On the discussion factor, ratings increase consistently as level increases.

As was seen with ratings of overall quality, the general trend for

ratings to be higher for higher level courses does not hold in some cases. Level 4 is rated lower than other levels on the interest/atmosphere and evaluation skills, and there are no differences in ratings of organization and clarity across levels. The administrator who is interpreting results or planning teaching improvement programs should consider these variations.

Class Size

An overall significant multivariate difference was found among the four categories of class sizes. When the factors were examined separately, the only significant difference among sizes appeared on the discussion factor, although organization and clarity approached significance. The difference found on the discussion factor followed the expected trend; that is, smaller classes received higher ratings. The most dramatic drop in ratings was observed for classes over 100 in size.

These results are particularly interesting in that it appears that the effect of class size only occurs for one teaching skill, that of discussion leading; however this effect is strong enough that it produces a general trend of differences in ratings. It may be the case that other teaching skills are relatively unaffected by the size of the classes.

Further Teaching Skills Analyses

Again, since differences were observed among departments, levels of instruction and class sizes, further analyses were performed in order to assess whether or not these differences were consistent.

Level of instruction within each department. The extent to which ratings varied among levels of instruction was found to be inconsistent across departments. In department 1 an overall significant multivariate difference

was found. When the factors were examined separately, the differences among levels existed only for the evaluation factor and for the discussion factor. On the evaluation factor, the variation was different from the expected trend: level 2 students were rating this skill highly, and level 6 students tended to give low ratings. Higher level students may become more demanding about evaluation procedures, or the nature of the learning taking place may be more difficult to evaluate. On the discussion factor, the differences were as expected. In department 2, an overall significant multivariate difference was found. However further analysis revealed that this difference only existed for the discussion factor. The expected trend was not followed exactly; level 2 students were rating discussion higher than were students at the other levels. In department 3 an overall significant difference was again found, and univariately it existed for the organization, evaluation and discussion factors. The general trend of higher ratings at higher levels was not followed. In each case, level 6 students were rating the teaching skills lower than students at the other levels. Ratings of the discussion factor consistently decreased as level increased, the inverse of the usual relationship. In department 4, no differences among levels of instruction were found. In department 5, an overall significant multivariate difference was found, and univariate differences were found on the interest/atmosphere, valuation and discussion factors. On the interest factor, level 4 ratings were lower than those at the other levels; on the evaluation factor, both level 4 and 6 were lower than would be expected, and the ratings of the discussion factor followed the expected trend, with the higher level students giving higher ratings.

As has been observed repeatedly in these analyses, individual departments often deviate from the usual relationships among course

characteristics and student ratings. Administrative users of evaluation results should consider this finding.

Class size within each department. When the overall effect of class size on the ratings of teaching skills was examined, it was found to exist only for the discussion factor, and to follow the expected trend of higher ratings for smaller classes. When examined within each department, the same results were generally found. Departments 1, 2 and 5 showed these results. In department 3, ratings of the interest factor were also affected by class size, and in a different direction. The largest class size received the highest ratings on interest. On the discussion factor, the usual trend was followed. In department 4, differences were found on the organization and clarity, evaluation and discussion factors. For the organization factor, a class size of 21 to 40 yielded the lowest rating; otherwise the trend was followed. Ratings of the evaluation factor were highest for the largest class (over 100 students); however the trend was followed for the remaining class sizes. Ratings of the discussion factor were as expected, with smaller classes giving higher ratings.

As mentioned earlier, the strongest effect of class size is on students' ratings of discussion-related items, and this holds true in all departments. There is, however, some variation among departments; that is, in some subject areas, class size also affects the ratings of other teaching skills, and in unexpected ways. Again, this should be considered in the interpretation of evaluation results.

III. Prediction of Overall Quality from Factors

The differences across departments, levels of instruction and class sizes has been examined for both ratings of overall quality and for clusters of specific teaching skills. The next issue that is of interest in the interpretation of these results is the extent to which ratings of teaching skills or clusters of teaching skills predict overall ratings of instruction. In other words, if a particular set of courses is rated highly, is this mostly due to a specific aspect of teaching? If so, then in order to improve the ratings of overall quality, one might then wish to concentrate on the improvement of that aspect of instruction. Since ratings have been shown to vary across department, level of instruction, and class size, it also becomes important to examine whether or not the relationships between aspects of teaching and overall ratings varies across these variables.

Regression analyses were conducted in which the four factor scores used in Question 2 were the independent variables predicting each of the overall ratings (amount learned, significance of learning, effectiveness of the instructor, and overall value).

Overall Relationship

The general trend was for the first factor, interest/atmosphere to account for the most variance in the prediction of overall ratings. The strongest relationship was found in the prediction of the effectiveness of the instructor, where this factor accounted for 49% of the variance, and the weakest relationship was with the significance of what was learned (20% of the variance accounted for). Factor 2 (organization and clarity) was consistently the next best predictor, accounting for 23% of the variance of instructor effectiveness, and only 7% for the significance of learning.

Factors 3 and 4 (evaluation, discussion) followed in that order. All predictions were statistically significant.

Department

When the analyses were repeated within each department, some deviations from this trend were found. In departments 1 and 5, factor 2 (organization and clarity) was the best predictor of amount learned, instructor effectiveness, and overall value, accounting for between 30% and 49% of the variance. Factor 2 (interest/atmosphere) remained the best predictor of significance of learning, although it only accounted for 21% and 14% of the variance. In general, department 1 was a more highly rated department and 5 was rated significantly lower in several cases.

The remaining departments tended to be consistent with the overall trend. That is, Factor 1 was always the best predictor, accounting for 56% to 21% of the variance. Factor 1 always accounted for the highest percentage of the variance in predicting instructor effectiveness, which would be expected. In some cases, Factors 2 and 3 (organization and clarity; evaluation) changed order in their prediction, although there were no striking differences in terms of variance accounted for.

In summary, within some subject areas, different teaching skills are more highly related to ratings of overall quality. It would be important, in teaching improvement efforts, to determine which aspects of instruction students most closely associate with an overall "good" course.

Level

When each of the levels of instruction was examined separately, some deviations from the general trend were found. For the lower levels (2 and 3) the second factor (organization and clarity) was the best predictor of some of the overall ratings. At level 2, factor 2 best predicted amount learned, instructor effectiveness, and overall value, accounting for 29% to 44% (for instructor effectiveness) of the variance. At level 3, the organization factor was the best predictor of amount learned, and overall value, but not either significance of learning or instructor effectiveness. Levels 4 and 6 followed the general trend more closely, although at level 4, the third factor (evaluation) was the second predictor for all the overall ratings except instructor effectiveness. This is consistent with Question 2 results, where Level 4 was significantly lower than the other levels on the ratings of the evaluation aspect of instruction.

Within levels, then, students seem to perceive different aspects of instruction as contributing to their overall ratings of the courses. Particularly at the lower levels, the organization and clarity of the instruction contributes more to overall effectiveness than it does in general. And at the fourth level, evaluation procedures are perceived to be more important than they are in general, although the interest and learning atmosphere aspect of the instruction is still the most closely related to overall quality.

Class Size

When the relationships among ratings of teaching skills and ratings of overall quality of instruction were examined separately for different class sizes, the most striking deviation from the overall trend occurred with the

largest classes (over 100 students). For the class sizes up to 100, the interest/atmosphere skill always accounted for the most variance in the predictions of overall quality (20% to 54% of the variance). In the smallest classes (less than 20) evaluation or discussion was the second best predictor, and organization was seen to be relatively unimportant. However for classes of 21 to 40 and 41 to 90, organization was the second best predictor of ratings of overall quality. In the large classes, ratings of the evaluation procedures were most closely related to amount learned, and discussion best predicted the significance of learning and the overall value, although it accounted for only 12% and 28% of the variance. The interest/atmosphere factor was only relevant in terms of overall instructor effectiveness (41% of the variance).

For class size, the only change in the tendency for interest/atmosphere to best predict overall quality was at the largest class size. This implies that if attempts are being made to improve instruction in large classes, this may be an area which need not be considered; rather organization and clarity are more likely to be relevant to students' perceptions of the "good" course.

Since the relationships among ratings of teaching skills or factors and ratings of overall quality varied somewhat dependent on subject area, level of instruction and class size, these results were examined further. Only exceptions to the relationships presented earlier will be discussed.

Level of instruction within each department. When factor scores were used to predict overall ratings for each level of instruction, the lower levels (2 and 3) tended to differ from the higher levels in each department, as was found when levels were examined over all departments. For department 1, the organization and clarity factor was the best predictor of the overall

ratings for levels 2 and 3, and the interest/atmosphere factor was the best predictor for the higher levels of instruction. In department 2 the best predictor at the lower levels varied dependent on the item: evaluation and discussion contributed to amount learned, and to the significance of what was learned; evaluation and organization contributed to instructor effectiveness; organization was the best predictor of overall value. Level 4 and 6 students perceived the interest/atmosphere skill to be the most closely related to all the overall quality ratings. In department 3, the interest/atmosphere factor was the best predictor of overall quality for all levels of instruction, and in department 4 only the lowest level deviated from this trend, with evaluation and organization (for instructor effectiveness) being seen as the most relevant factors. In department 5, organization and clarity tended to remain the best predictor of overall quality. At level 2, the interest/atmosphere factor was perceived to be of the most importance in terms of significance of learning, and at level 4, the interest/atmosphere factor was the best predictor of amount learned (55% of the variance), significance of learning (44%) and instructor effectiveness (49%). At the graduate level, evaluation procedures became the best predictor for all items except the significance of what was leraned (where organization and clarity were seen to be the most relevant).

In general, then, students in different levels in each department tend to consider different specific teaching skills as most closely related to the overall quality of instruction. The organization and the evaluation skills are most often more relevant at the lower levels, although in department 5, which is more quantitative in terms of the subject area, these skills were

closely related to overall quality at the higher levels.

Class size within each department. When the relationship of ratings of teaching skills and ratings of the overall quality of instruction are examined within individual departments, the trend was similar to that found in the investigation of class sizes regardless of department. That is, for the smaller classes, the interest/atmosphere factor was the best predictor of overall quality, but this changed in the larger classes, where organization and evaluation procedures became more relevant. In department 1 for example, the smallest classes (1-20) always saw interest/atmosphere skills as most closely related to overall quality. For classes of size 21-40, organization was most relevant in terms of amount learned and the overall value of the course. In classes of size 41 to 99, organization was the best predictor of all ratings except the significance of learning (where interest/atmosphere was still the best) and in the largest classes organization was always the most closely related to overall quality (accounting for 48% to 66% of the variance). The pattern for the different class sizes was similar in each department and need not be described individually.

IV. Personnel Decisions

The fourth question asked of the data was concerned with the factors that should be considered in the making of personnel decisions. Although no additional analyses were done, results from questions 1 to 3 provide useful information for the administrator who is considering making comparisons among instructors on student questionnaire results.

First, subject area, level of instruction and class size were found to influence student ratings. When these variables are examined separately, there appear to be fairly straightforward relationships between different

types of courses and the student ratings. However, further analyses of the interactions among the course characteristics revealed complex relationships which do not easily allow the establishing of norms for comparisons. Depending on the particular department or level being examined, the trends usually reported in the literature (and those found here in the overall analyses) do not necessarily hold true. The administrator should, ideally, examine the nature of these relationships for the particular group of faculty about whom decisions are being made. If this is not possible, or desirable, then professors could be compared to others teaching the same or similar courses, or absolute standards could be set by those faculty involved. It must be emphasized that given the variation among ratings across types of courses, and the complexity of the interactions, comparisons among instructors should be made with extreme caution.

Secondly, it was found that ratings of teaching skills vary in complex ways dependent on course characteristics. Also, students in different disciplines and at different levels of instruction perceive different relationships among these teaching skills and the overall quality of the instruction. That is, students in one discipline or at one level may see organization and clarity as being the teaching skill most closely related to overall effectiveness, whereas another group of students may emphasize interest, or evaluation procedures. Therefore, when personnel decisions are being made, it would be dangerous to consider all teaching skills equally. It might be the case that an individual instructor who is rated highly on, for example, organization, but not on other skills, is, in fact, providing a particular group of students with effective instruction.

As can be seen from these results, the use of student questionnaire data in personnel decision-making is not a straightforward issue. Great care

must be taken in order to be fair to individual instructors, and other sources of evaluation information should be included in order to conduct a comprehensive evaluation.

V. Curriculum Decisions

The final question asked of the data was whether or not specific courses were rated significantly differently from each other regardless of the year in which they were taught, or the instructor. For each department, courses that were taught more than ten times were used as levels of the independent variable in one-way multivariate analyses of variance. The four overall ratings of quality were the four dependent variables.

Overall significant multivariate differences were found for each department; these differences do not appear to be due to level or class size, since generally the courses receiving the highest and the lowest ratings in a department were at the same or similar levels. These results will be summarized briefly. The details of such results would only be of interest to an administrator in the relevant department or faculty; however the procedure is one which could be repeated in any institution where such data have been accumulated.

In department 1 a significant multivariate difference was found among the thirteen courses that had been taught more than ten times. Univariately, this difference existed for all four overall ratings. On three of the ratings, the courses that received the highest ratings and the courses that received the lowest ratings were at the fourth level. In the rating of instructor effectiveness, the lowest ratings were given in a third level course. In department 2, a significant multivariate difference was found; however further analysis showed that this was due only to a difference

among courses on the "significance of what was learned" item. Out of the eight courses included in the analysis, the highest rating was given in a level 6 class and the lowest in a level 4 class. In department 3 a significant multivariate difference was found, and univariately, the fifteen courses differed on all four overall ratings. Again both the highest and the lowest ratings tended to be given in level 4 courses, with the exception of a low rating at level 2 on "significance of what was learned" and a low rating at the third level on instructor effectiveness. A significant multivariate, and four significant univariate differences were found in department 4. For each overall rating, both the highest and the lowest means were found in the same two level 4 courses. In department 5, a significant multivariate difference was found, but univariately, it was only due to ratings of amount learned and instructor effectiveness. For amount learned, the highest rating was given to a level 2 course and the lowest to a level 4 course; for instructor effectiveness, the highest rating was given to a level 3 course and, again, the lowest to a level 4 course.

In summary, this type of analysis indicates that there are specific courses within departments that are consistently rated lower or higher than others. A department chairman or a curriculum committee could utilize this information and select courses for revision.

CONCLUSIONS

Although previous research which has examined the relationships among course characteristics and student ratings of instruction has tended to yield rather straightforward results, it has been found that when these issues are investigated in detail, there are many cases where the general trends do not

hold. The purpose of this study was to generate and attempt to answer questions which a university administrator might ask of cumulated questionnaire data. As such, the answer, on the most general level, is that complex relationships exist among course characteristics such as subject area, level of instruction, class size, and students' overall ratings of instruction and of teaching skills.

Given a data base such as the one used for this investigation, it is possible to answer questions of the following type: (a) overall, which departments or subject areas receive the lowest and highest student ratings; (b) in general how do class size, level of instruction and any other variables of interest influence student ratings of the overall quality of instruction; (c) within any one department, or at any one level of instruction are these influences the same; (d) when ratings of teaching skills are examined, as opposed to overall ratings what is the nature of the relationships listed in the above questions; (e) how are ratings of teaching skills related to overall ratings, i.e., what teaching skills do students see to be most important in terms of the overall quality of instruction; (f) does the relationship between ratings of teaching skills and of overall quality vary dependent on subject area, level of instruction, class size, etc., i.e., do different groups of students value different teaching skills to different degrees; and (g) are there courses which are consistently rated lower or higher than others, regardless of the instructors or the time when they are taught.

For this particular set of data, it was found that: (a) departments differed significantly on overall ratings of instruction; (b) class size and level of instruction affected all overall ratings except those of instructor effectiveness; (c) the effect of class size and level of instruction varied dramatically across departments; (d) ratings of teaching skills were different

across departments, levels of instruction and class sizes, and within departments these relationships varied further; (e) overall, the interest/atmosphere factor was the best predictor of overall ratings of instruction; (f) when this relationship was examined in different subgroupings, it tended to vary; specifically, lower level students in general placed more value on organization and clarity; and (g) it is possible to pinpoint courses which are generally rated lower or higher than others, and this does not seem to be dependent on level: both the highest and the lowest ratings were most often received by level 4 courses.

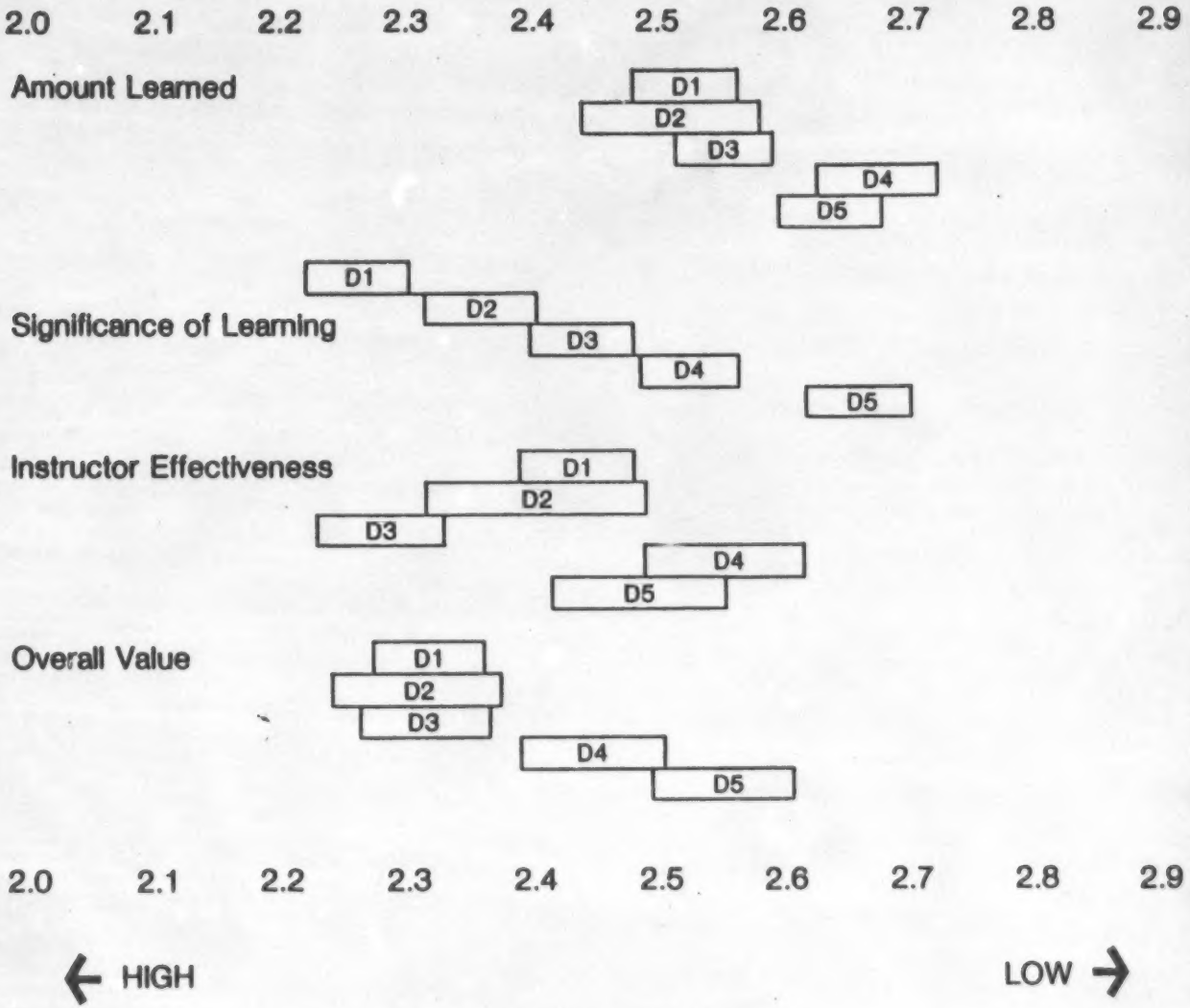


Figure 1. Overall Quality by Department

Legend:
D1: Accounting
D2: Finance
D3: Management
D4: Marketing
D5: Quantitative Methods

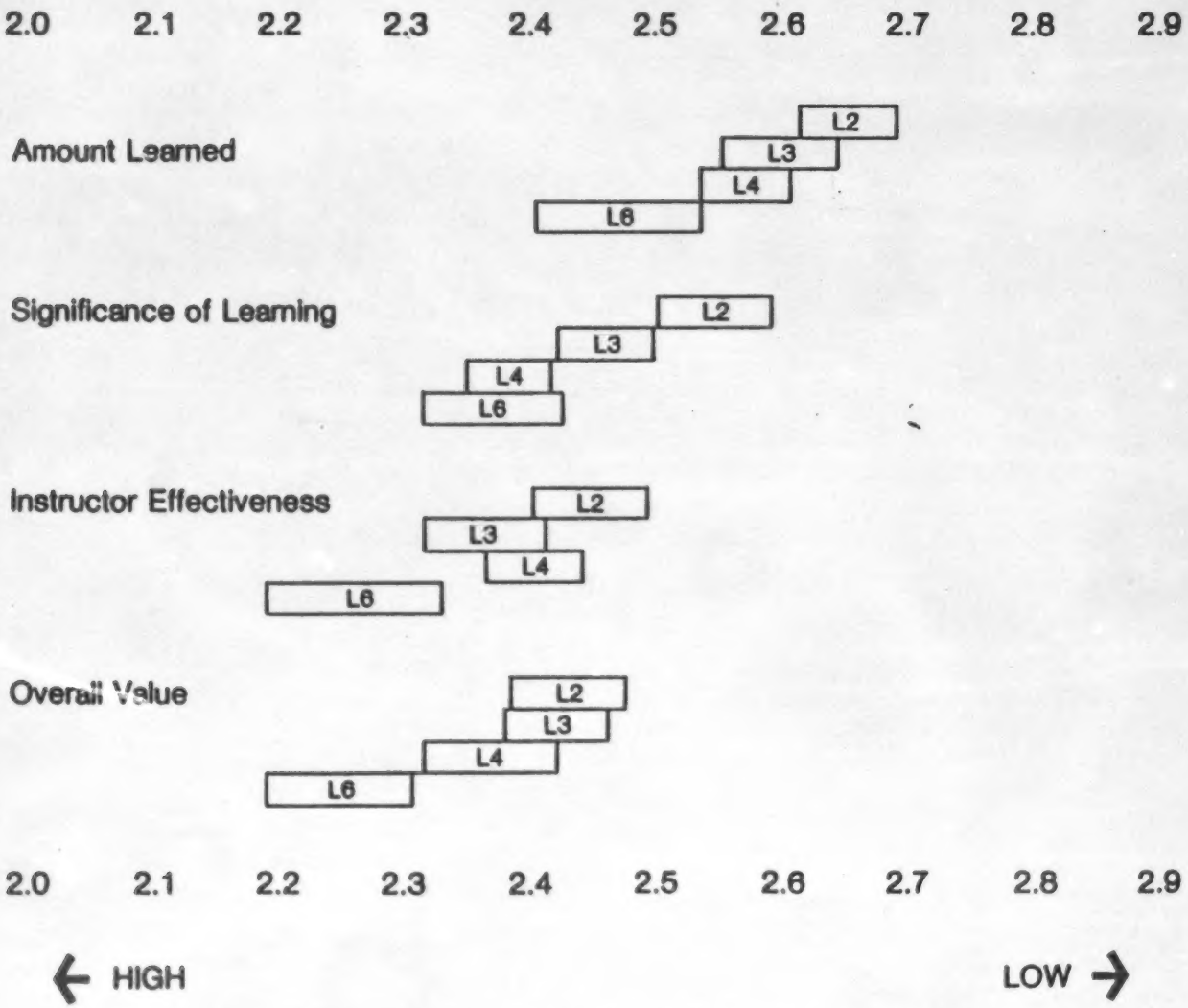


Figure2. Overall Quality by Level of Instruction

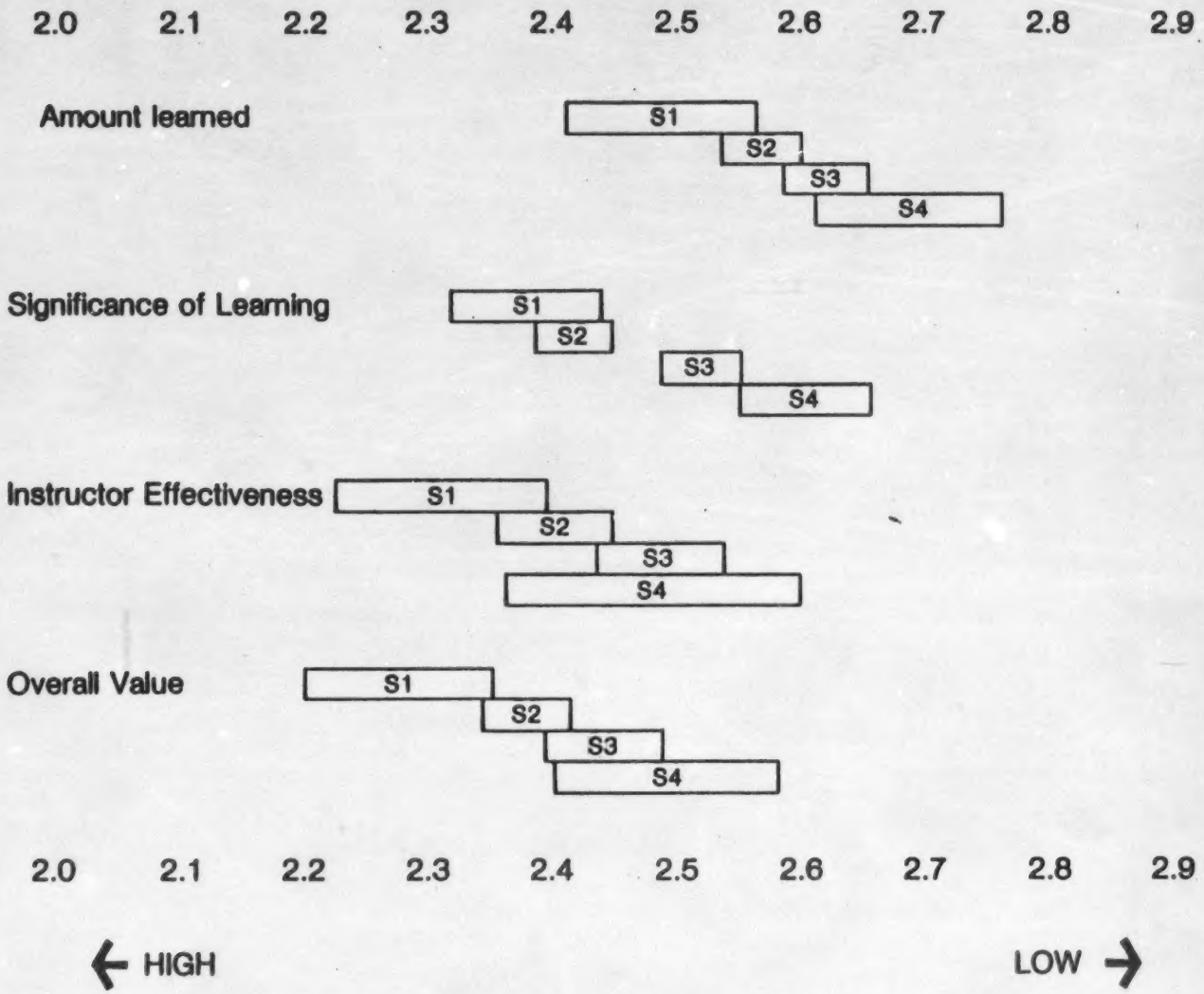
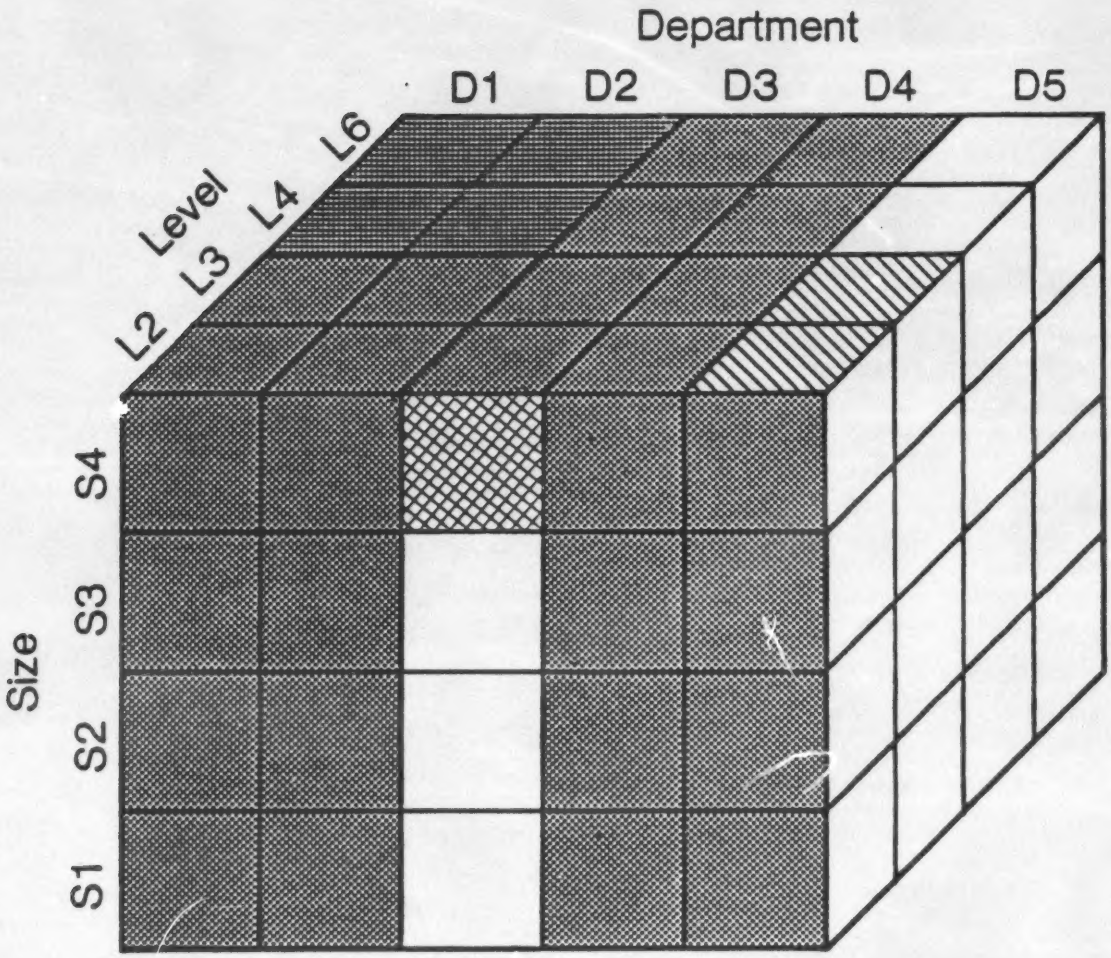


Figure 3. Overall Quality by Class Size

Factor: Interest



GENERAL TRENDS: higher ratings with higher levels
higher ratings with smaller classes

Follows trend

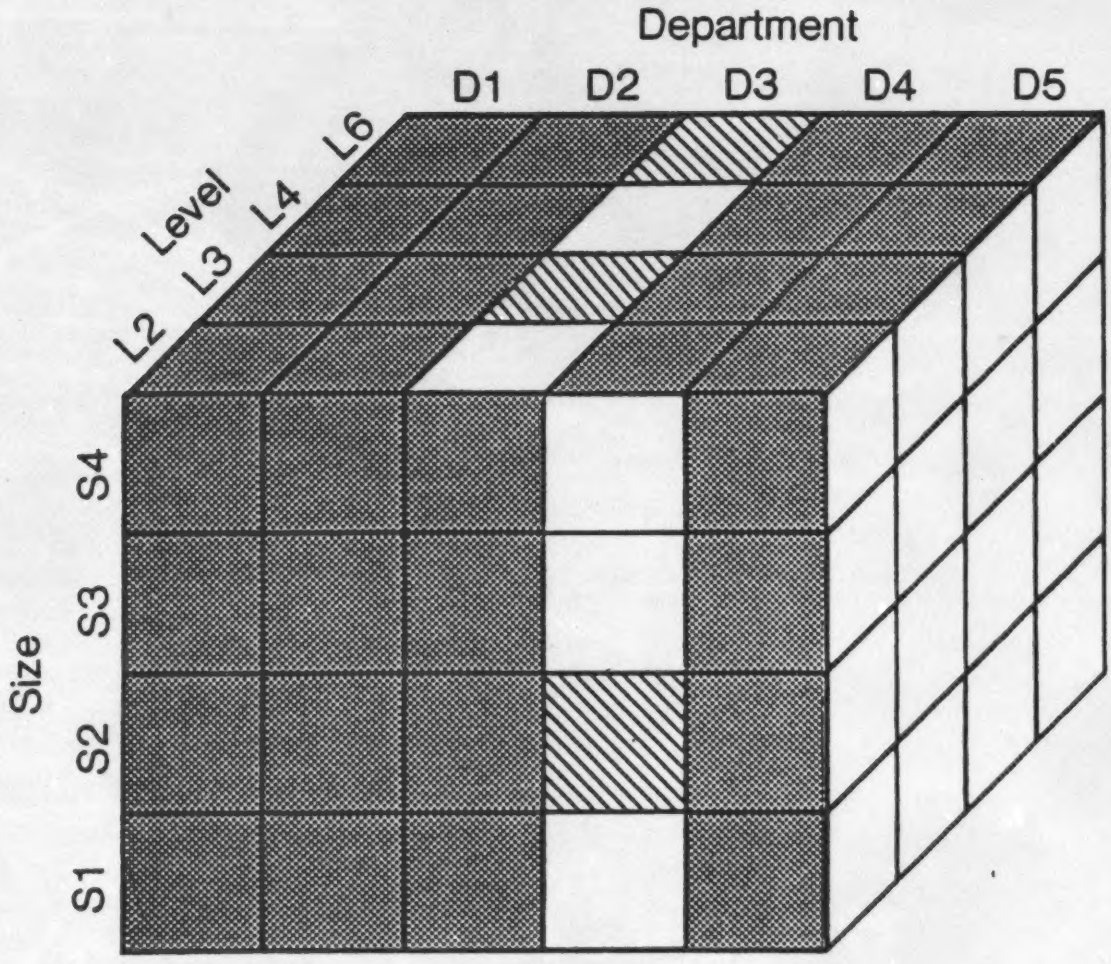
No difference

Unexpected high rating

Unexpected low rating

Figure 4

Factor: Organization



GENERAL TRENDS: higher ratings with higher levels
higher ratings with smaller classes

Follows trend

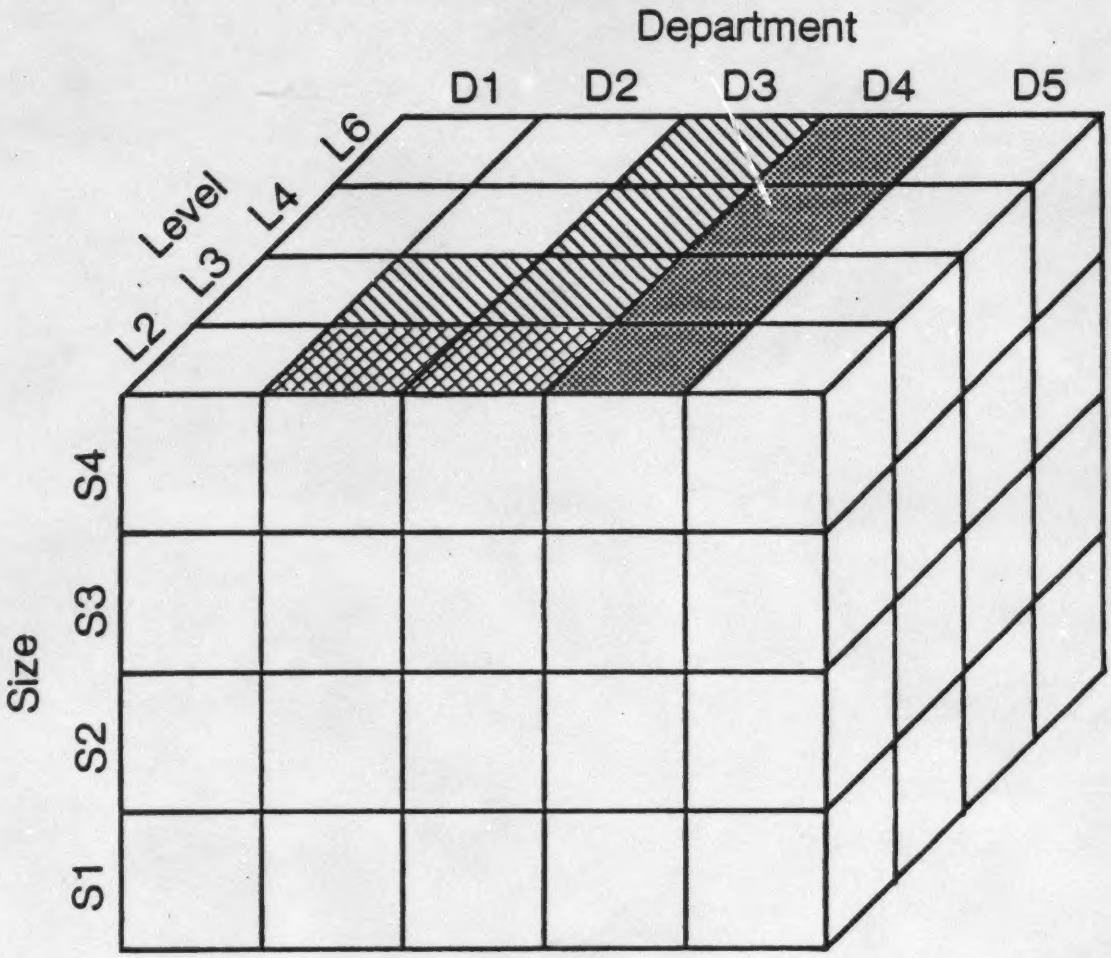
No difference

Unexpected high rating

Unexpected low rating

Figure 5

Factor: Discussion



GENERAL TRENDS: higher ratings with higher levels
higher ratings with smaller classes

Follows trend

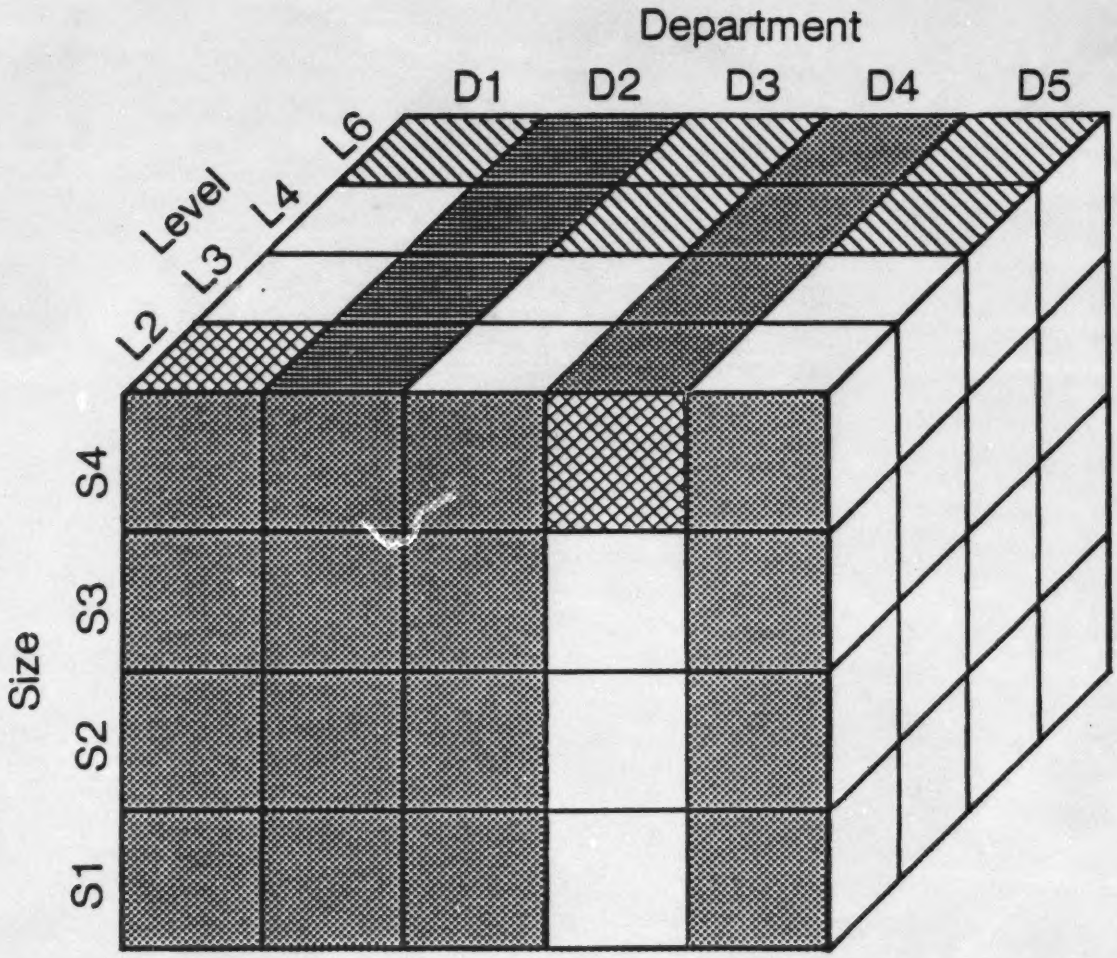
No difference

Unexpected high rating

Unexpected low rating

Figure 6

Factor: Evaluation



GENERAL TRENDS: higher ratings with higher levels
higher ratings with smaller classes

Follows trend

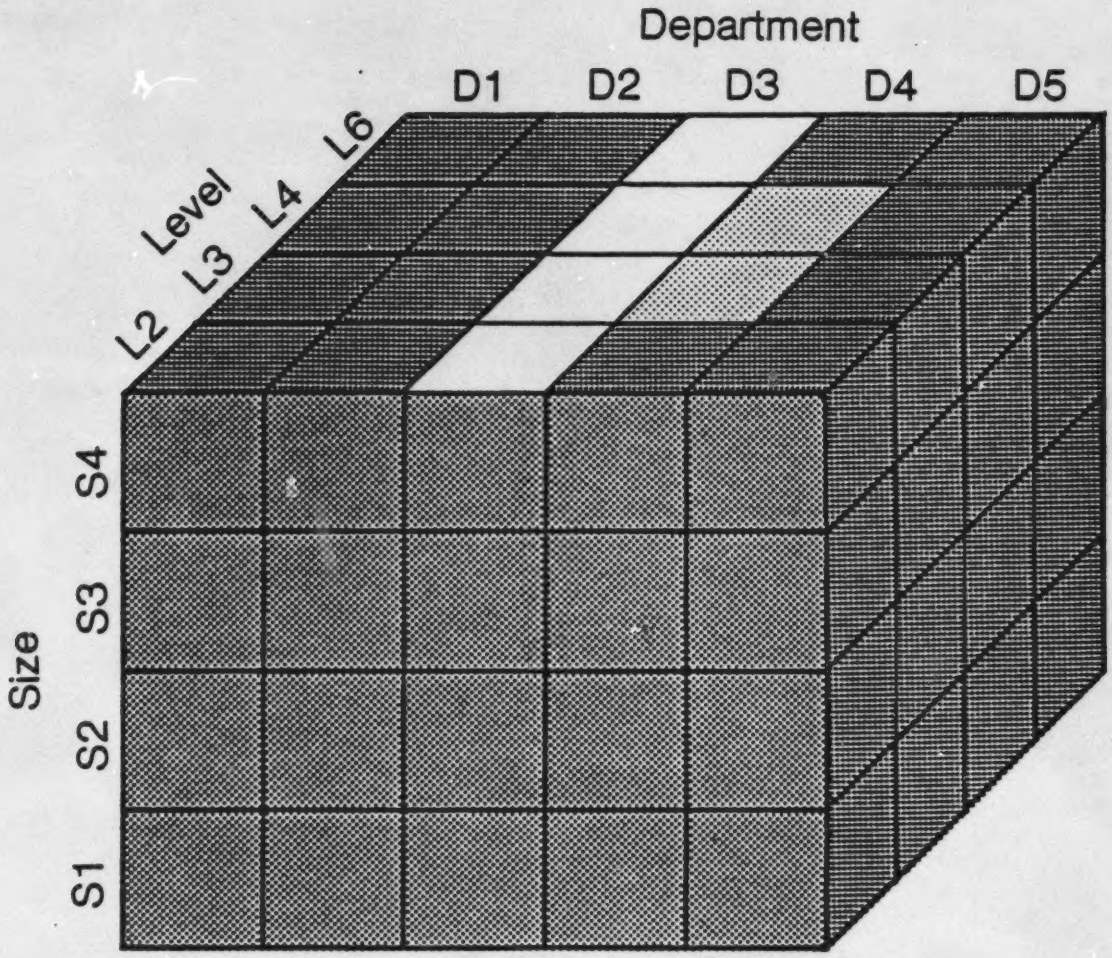
No difference

Unexpected high rating

Unexpected low rating

Figure 7

Item: How much you are learning



GENERAL TRENDS: higher ratings with higher levels
higher ratings with smaller classes

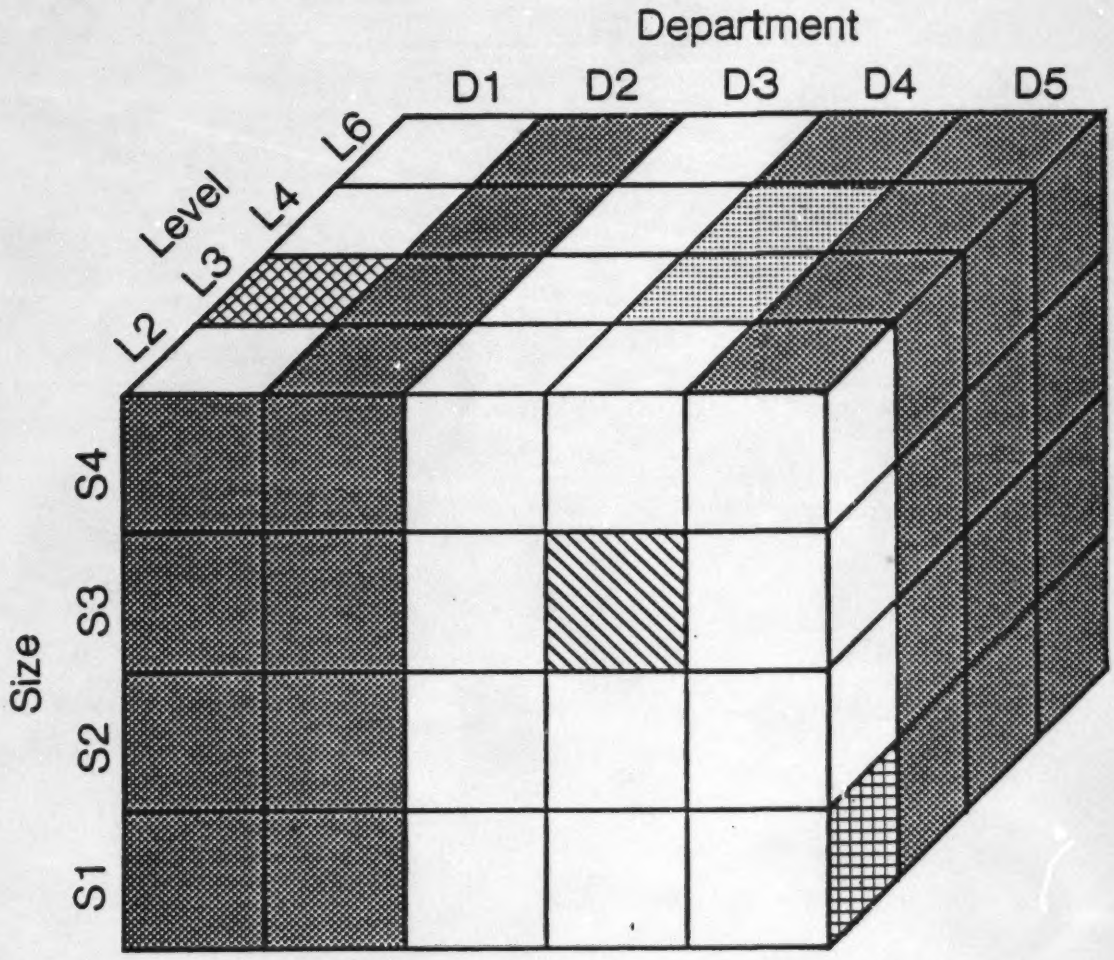
Follows trend

No difference

Difference in pattern

Figure 8

Item: Significance of what you are asked to learn



GENERAL TRENDS: higher ratings with higher levels
higher ratings with smaller classes

Follows trend

Unexpected high rating

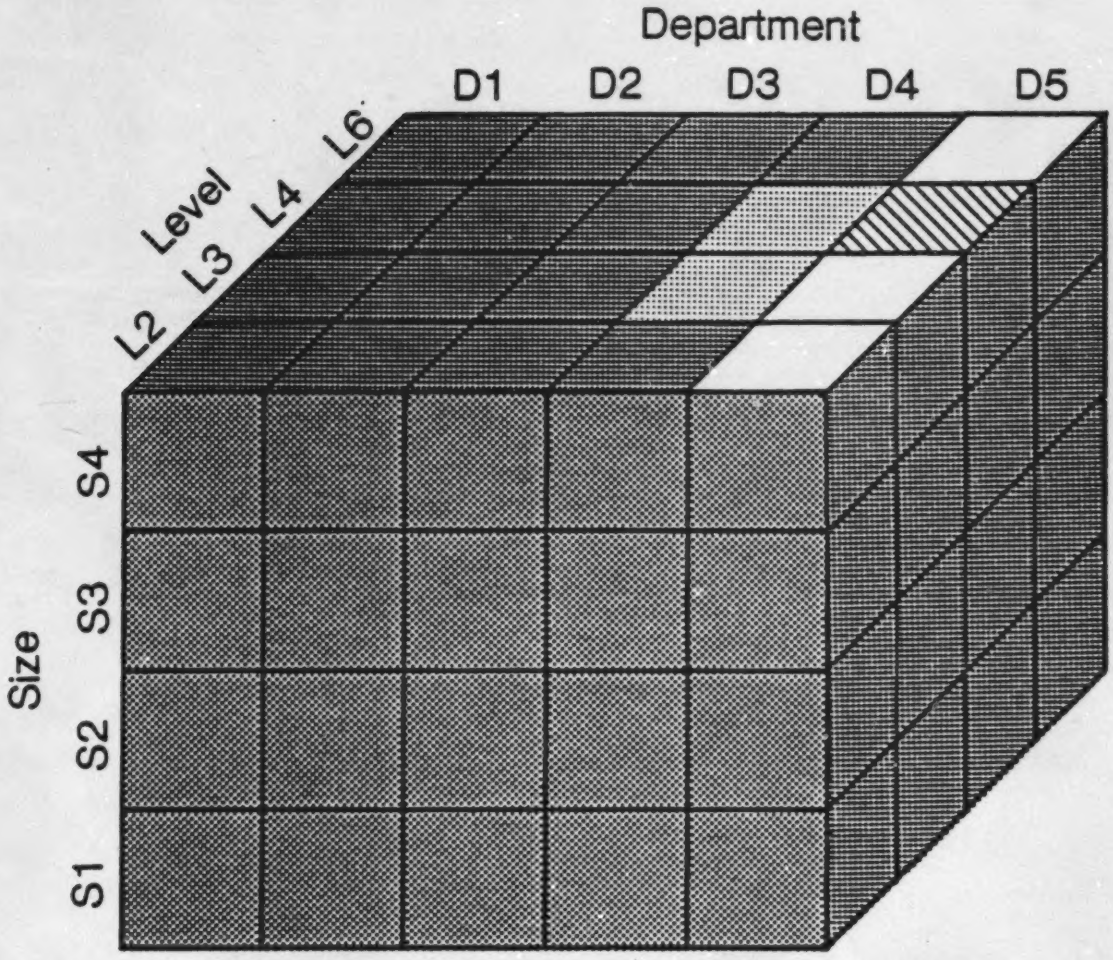
Unexpected low rating

No difference

Unexpected difference in pattern

Figure 9

Item: Effectiveness of the Instructor



GENERAL TRENDS: higher ratings with higher levels
higher ratings with smaller classes

Follows trend

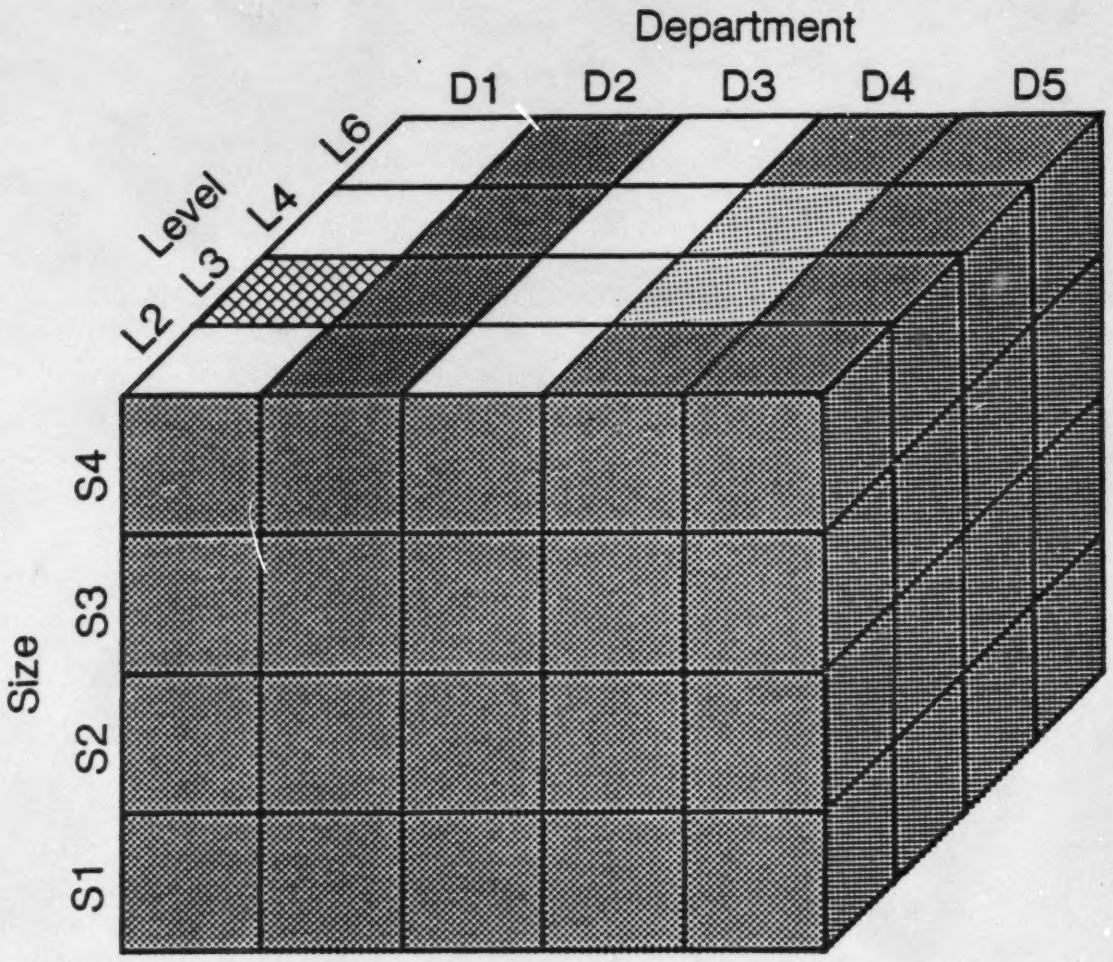
No difference

Difference in pattern

Unexpected low rating

Figure 10

Item: Overall Value of the Course



GENERAL TRENDS: higher ratings with higher levels
higher ratings with smaller classes

Follows trend

No difference

Difference in pattern

Unexpected high rating

Figure 11